Abstract of the Disclosure

A system, method, and processor readable medium for determining a language in which a document is created. After receiving an electronic document, an appropriate character set encoding (or encodings) for the text of the electronic document is determined. The character set encoding(s) indicate a list of potential languages in which the electronic document is created. The potential languages may be identified using bit flags. The number of potential languages for which an electronic document is created may be increased or decreased according to predetermined criteria. The number of potential languages may be adjusted by comparing groups of characters (n-grams) included in the electronic document with entries in a look-up table. If n-grams are located in the look-up table, bit flags associated with the n-grams may be logically ANDed together. This process may be repeated until only a single bit flag remains. The remaining bit flag identifies the language in which the electronic document is created.

RES 56988v2